What is causal about individual differences? : A comment on Weinberger

Article *in* Theory & Psychology · June 2015

DOI: 10.1177/0959354315587784

1 author:

Denny Borsboom
University of Amsterdam
166 PUBLICATIONS **4,126**
CITATIONS

Comment

Theory & Psychology

Theory & Psychology
1–7

DOI: 10.1177/0959354315587784
tap.sagepub.com
SAGE

What is causal about individual differences? : A comment on Weinberger

Denny Borsboom
University of Amsterdam

Abstract

Weinberger (2015) claims that if a latent variable is a cause, it must be a within-subject cause. In addition, Weinberger suggests that this fact refutes the conclusion of Borsboom, Mellenbergh, and Van Heerden (2003), who stated that standard psychometric models have a causal interpretation that is cast strictly in a between-subjects sense: individual differences in the latent variable may cause individual differences in test scores, while the latent variable has no causal relevance at the level of the individual. Weinberger's argument elucidates the status of causal relations in latent variable models, and clearly spells out the strong assumptions that underlie the use of such models. However, contrary to Weinberger's claims, a pure individual-differences reading of the causal model is possible. This interpretation relies on the fact that, for latent variable models, shifts of the person relative to the latent dimension can either be interpreted as a change of the individual, or as a shift of the population relative to the individual. The latter interpretation does not require us to place assumptions on what interventions would do intra-individually, but nevertheless is consistent with a causal interpretation along the lines suggested by Weinberger.

Keywords

causality, individual differences, latent variables, philosophy of science, psychometrics

Weinberger (2015) disputes the claim, formulated in Borsboom, Mellenbergh, and Van Heerden (2003), that standard psychometric measurement models can be given a causal interpretation that is exclusively confined to individual differences but at the same time remains consistent with modern theories of causality, including the causal calculus based on the do-operator, as proposed by Pearl (2009). He argues that, within Pearl's (2009)

Corresponding author:
Denny Borsboom, Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA, Amsterdam, The Netherlands.
Email: d.borsboom@uva.nl

framework, if a variable is a cause in a population, it must be a cause in at least one subpopulation, and from this he infers that the latent variable in question must support counterfactual claims at the level of at least one individual element of the population, which Weinberger takes to be the individual person.[1]

Thus, Weinberger argues that if we claim that, say, "variation in the *g*-factor of intelligence causes variation in IQ-test scores," we are, as a matter of philosophical necessity, committed to counterfactuals of the form "if we changed John's intelligence through an intervention, then his test scores would change in accordance with the *g*-factor model." This implies that between-subjects variation and within-subject variation have the same cause (Adolf, Schuurman, Borkenau, Borsboom, & Dolan, 2014; Kievit, Frankenhuis, Waldorp, & Borsboom, 2013), which would in turn mean that, of the model interpretations discussed in Borsboom et al. (2003), only the locally homogeneous case would sustain a causal analysis within Pearl's (2009) theory.

Weinberger's paper (2015) offers a detailed exposition of the problems exposed in Borsboom et al. (2003), connects the psychometric measurement model to the modern causal calculus in greater detail, and also corrects a number of inaccuracies in our paper. Further, from a pragmatic perspective, there is much to be said for the plausibility of Weinberger's conclusion. However, I am not convinced that the argument Weinberger gives is watertight. Even within Pearl's (2009) calculus, it appears to me that a between-subjects interpretation of the latent variable model, which is nevertheless causal, cannot be decisively ruled out, as would be required for Weinberger's conclusion to hold.

Causal effects and individual differences

To get a better understanding of what kind of causal relations may be formulated within a psychometric model, it is important to note that latent variables express only where a person is located relative to the rest of the population. In this sense, positions of individual persons on a psychometric latent variable are comparable to, say, IQ-scores, which are scaled to have a population mean of 100 and a standard deviation of 15. Clearly, in the case of IQ, a score of 115 means nothing more than that the relevant person scores one standard deviation above the population mean. To assess whether, say, John's IQ-score is "really" 115, it will not help to open up his brain, scan his genome, assess his use of problem-solving strategies, or measure the speed of his nerve fibers. The only thing that fixes John's IQ-score is where he is located relative to the rest of the population.

The same thing holds for latent variables. Positions of individuals on the latent variable are not parameters in the model that can be estimated or meaningfully specified, even in the most restrictive models (e.g., Rasch, 1960); all that one can specify, estimate, or determine are *differences between* positions of individuals (at best, one can estimate the metric distances between these positions, but typically an ordering is more realistic). Thus, latent variables provide ordinal or at best interval representations of the individual differences under study (Ellis & Junker, 1997). They are not absolute or ratio scales, for which the actual values of a numerical representation have a direct empirical interpretation independent of the individual differences between measured entities (Krantz, Luce, Suppes, & Tversky, 1971), nor are such scales assumed in the background.

This means that in a typical *g*-factor model for intelligence, which is, say, scaled to a latent normal density function with zero mean and unit variance, the truth condition of the statement "John's *g*-factor score is 1" is simply that 84% of the individuals in the population have a value below John's, and nothing more. The truth conditions of statements about individuals' positions on the latent variable are thus exactly parallel to those of IQ-scores. However, *unlike* IQ scores (which are transformations of "actual" raw scores, i.e., of the number of items correctly answered) latent variable positions are *not* transformations of anything. So, in the model formulation, there is nothing "beyond" the individual differences that may identify a person's position on the latent variable, independent of the population (one can of course fantasize about what may lie beyond the individual differences, but then one goes quite far beyond the actual empirical content of the model). Thus, while John's raw total score of correctly answered items on an IQ-test may be, say, 40 items, independent of what anybody else did, there is no parallel value or score on the latent variable that underlies his relative position on that latent variable.

Now assume that, say, a factor model is true in the population. This means that the probability distributions of the observed psychometric scores (e.g., subtest scores on an IQ-test) follow a multivariate normal density, which can be represented as a function of a single latent variable. Denoting individual *i*'s score on test *j* as a random variable X_{ij}, the model then says that the expectation of this random variable is $E(X_{ij} |\Theta=\theta_i)=\lambda_j\theta_i$, in which λ_j is the factor loading of test *j* and θ_i is *i*'s position on the latent variable (relative to the rest of the population). As Weinberger correctly notes, the causal interpretation of this model follows the same rules as the causal interpretation of any regression model. This means that if we don't just observe the value $\Theta=\theta_i$ but instead *intervene* on Θ to take the value θ_i (this is expressed as $do(\Theta=\theta_i)$) then we should get the same result: $E(X_{ij} |\Theta=\theta_i)= E(X_{ij} \mid do(\Theta=\theta_i))$. So far so good.

Weinberger's trouble starts when he begins to interpret the statement $do(\Theta=\theta_i)$. He likens this to an experimental situation, in which we study the effect of smoking (denoted as the variable S, with S=1 for smokers and S=0 for non-smokers). It is important to note, however, that S is *not* an individual differences-variable like the *g*-factor or the latent variables of psychometrics. That is, the truth condition for the statement "John's value on the variable S equals 1" does *not* involve a comparison to other people, but is simply *that John smokes*. It does not matter one bit whether John is one of many smokers, the only smoker in the population, or, for that matter, the only person left alive on Earth. Smoking is an attribute that is present intrinsically at the level of the individual, and that is directly coded into the variable S. For this reason, $do(S=1)$ can only mean that an intervention forces a person to smoke, and, likewise, $do(S=0)$ can only mean that an intervention forces a person not to smoke. As a result, a purely between-subjects interpretation of the causal effect of smoking on, say, lung cancer is out of the question. This is not because we *can* intervene intra-individually on S, but because *any* intervention that changes S is, by necessity, intra-individual in nature.

However, $do(\Theta=\theta_i)$ is not one bit like this, as may be intuitively grasped from the fact that, contrary to the variable S, for Θ we have no way (not as a matter of practical limitations, but *in principle*) to even determine whether John *has* the value $\Theta=\theta_i$ without comparing him to the rest of the people around. And this is where the problem comes from: Pearl's *do*-operator, as applied to the *g*-factor of intelligence, says *what would happen if*

we fixed the g-factor to a certain value. But this does not logically equate to the evaluation of *what would happen if we surgically intervened on John's intelligence*. The reason for this is simply that John's position on the *g*-factor is not a fact about John in isolation, but an expression of his relation to other individuals in some population. Thus, although John's position on the *g*-factor *may* be changeable by some intervention on *his* intelligence (say, shrinking his cerebral cortex or feeding him intelligence-boosting pills), unlike the case with smoking, this is *not* the only option available to us. The reason is that we can achieve the same result by manipulating the individual differences and leaving John to his own devices.

To make matters concrete, suppose that John has score 1 on the *g*-factor, and we consider the operation $do(g_{John} = 0)$. This operation can be realized by decreasing John's level on the *g*-factor, resulting in the new value g^*_{John}, such that $g^*_{John} = g_{John} - 1 = 0$; perhaps by manipulating "whatever realizes John's intelligence" as Weinberger optimistically supposes would be possible. However, we can also achieve our goal by changing the *g*-factor scores of other individuals, or by changing the composition of the population such that the population mean μ rises from zero to one while its standard deviation σ remains the same; in either case, John's new score on the *g*-factor also becomes $g^*_{John} = (1 - \mu)/\sigma = 0$. The model does not care about the distinction between these options, because the *meaning* of a score of 0 is *exhausted* by the relative position of the individual in the population, so any manipulation that changes that position is a legitimate intervention. I assume many such manipulations could be imagined, and they would all be equally valid.

In both of the cases considered above, we have changed John's position on the latent variable from 1 to 0. In the first case, Weinberger's conclusion holds: the model implies a counterfactual at the level of John's intelligence (or whatever realizes it). However, in the second case, we have also achieved our desired result, but *without* intervening at the level of John. One does, of course, have to express the test score or item response variable in a relative scale too (e.g., in terms of a Z-score) for the counterfactual to work, but this does not pose a problem: in the situations discussed in Borsboom et al. (2003), models can be fully standardized without loss of generality. This is in fact one of the spoiler-alerts in latent variable modeling: if such models really contained testable hypotheses at the level of the individual, it should not be possible to standardize them to the population distribution without losing some empirical implication somewhere along the road.

Perhaps Weinberger would counter that this is not a legitimate causal intervention, but it is not clear to me what the grounds for such an ordeal would be. As far as I can see, the intervention is a proper candidate for the application of the *do*-operator, and would respect all the standard model-based implications of latent variable models, like local independence, vanishing tetrads (Bollen & Ting, 1993), mediation-based proportionality constrains (Franić et al., 2013), and, if the population has additional causal structure, the implications of the mixture models Weinberger formulates in defense of his thesis. Thus, there exists at least one intervention that respects both the latent variable model and the modern causality calculus, but does not involve counterfactual surgeries at the level of the individual's psychological constitution.

It is useful to pause a bit on the nature of the individual-differences counterfactuals suggested above. Perhaps these may strike some as an academic curiosity. I do not think

they are. Instead, I submit that this type of causal claim is very common and arises in essentially all situations that ascribe causal force to positions of individuals in rank orders. The reason is that these positions are essentially relative, and always allow at least two counterfactuals at the individual level: one in which the individual is changed, and one in which the rest of the population is changed. We can say that finishing first caused Vincenzo Nibali to obtain the yellow jersey in the 2014 Tour de France, and accept the implied counterfactual that "if Nibali had finished second, he would not have worn the yellow jersey." However, the antecedent of this counterfactual ("Nibali finished second") need not involve changes in Nibali's actual race; we may alternatively imagine that someone else cycled faster. And although there certainly is a trivial sense in which this counterfactual can be interpreted at the level of Nibali (after all, the statement is *about* him), that does not appear to be the sense required for a within-person causal reading of interventions as imagined by Weinberger, and it is certainly not the kind of within-person interpretation that psychologists have in mind when they talk about the *g*-factor as a problem-solving module inside the head (Kanazawa, 2004) or about extraversion as a cause of going to parties (McCrae & Costa, 2008).

Thus, to revisit our example, if everyone else on the planet died, the counterfactual "if John had had a lower position on the *g*-factor, he would have obtained a lower IQ-score" would not be true anymore. Instead, it would instantly become devoid of empirical content, because the term "*g*-factor" in the antecedent, which designates a structure of individual differences, would become referentially defective without the individual differences that constitute it. This, I think, goes to the heart of the matter. The problem with the intra-individual interpretation of psychometric variables is not so much that they cannot have causal implications at the individual level (they clearly can, in a relative way), but in the fact that psychometric latent variables designate structures that typically do not exist at the level of the individual person.

Conclusion

Psychometric measurement models, as discussed in Borsboom et al. (2003), relate sets of individual differences to each other. That is all they do. Any interpretation that is given to the model, whether causal or not, should therefore be translatable in terms of individual differences. Weinberger's intra-individual causal models have tractable implications for the structure of individual differences, and thus have this property; however, they do not exhaust the space of possibilities. As this commentary shows, even within the restrictive framework of Pearl (2009), a consistent causal model can also be defined without making direct reference to intra-individual causation.[2] Thus, Weinberger's argument appears less than watertight.

Although in this comment I have given an example of how one could set up a causal account in terms of individual differences, that does not mean that I find the resulting model interpretation desirable, plausible, or sensible. As Borsboom et al. (2003) conclude, the pure between-subjects interpretation of latent variable models is unsatisfying and, to my mind, cannot support standard psychometric practices in psychology; for instance, those surrounding the *g*-factor of general intelligence (Jensen, 1999), the Big Five traits of personality (McCrae & Oliver, 1992), and the internalizing and

externalizing dimensions of psychopathology research (Caspi et al., 2014). When people search for "genes for intelligence," or study "the effect of antidepressants on depression," or want to know the "neural basis of personality," these attempts hardly make sense when interpreted as being purely directed at individual differences (see also Borsboom, 2005; Markus & Borsboom, 2013). However, the interpretation of latent variable models in terms of intra-individual processes, mechanisms, or properties is not logically forced upon us, even if we interpret the model causally.

Acknowledgements

I would like to thank Mijke Rhemtulla and Riet van Bork for their comments on an earlier draft of this paper.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. I do not think that this inference follows, as the subscript *i* in psychometric models need not refer to a person; it may also indicate a subpopulation (Holland, 1990). However, in view of space limitations, I will not pursue this issue here.
2. Of course, within more relaxed frameworks, such as the classic Millean paradigm (Mill, 1843) or probability-raising accounts (Suppes, 1970), it is considerably more straightforward to define causal interpretations that do not rely on intra-individual causal claims. However, I have not addressed such alternatives because, as Weinberger himself notes, his argument need not hold in accounts of causality other than Pearl's (2009).

References

Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: A distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology*, *5*, 883. doi:10.3389/fpsyg.2014.00883

Bollen, K. A., & Ting, K. (1993). Confirmatory tetrad analysis. *Sociological Methodology*, *23*, 147–175.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, UK: Cambridge University Press.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H. L., Israel, S. … Moffit, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*, 119–137.

Ellis, J. L., & Junker, B. W. (1997). Tail-measurability in monotone latent variable models. *Psychometrika*, *62*, 495–523.

Franić, S., Dolan, C. V., Borsboom, D., Hudziak, J. J., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2013). Can genetics help psychometrics? Improving dimensionality assessment through genetic factor modeling. *Psychological Methods*, *18*(3), 406–433. doi:10.1037/a0032755

Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, *55*, 577–601.
Jensen, A. R. (1999). *The g factor: The science of mental ability*. Westport, CT: Praeger.
Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, *111*, 512–523.
Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*, 513. doi:10.3389/fpsyg.2013.00513
Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. I). New York, NY: Academic Press.
Markus, K., & Borsboom, D. (2013). *Frontiers of validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.
McCrae, R. R., & Costa, P. T. Jr. (2008). Empirical and theoretical status of the Five-Factor Model of personality traits. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *SAGE handbook of personality theory and assessment*: *Vol. 1* (pp. 273–294). Los Angeles, CA: Sage.
McCrae, R. R., & Oliver, P. J. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality*, *60*(2), 175–215.
Mill, J. S. (1843). *A system of logic*. London, UK: Oxford University Press.
Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge, UK: Cambridge University Press.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Paedagogiske Institut.
Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam, the Netherlands: North-Holland.
Weinberger, N. (2015). If intelligence is a cause, it is a within-subjects cause. *Theory & Psychology*, *25*, XXX–XXX. doi:10.1177/0959354315569832

Author biography

Denny Borsboom is Professor at the University of Amsterdam. His work has mostly focused on philosophical and conceptual problems that arise from the application of psychometric models in psychology; examples concern the theoretical status of latent variables, the concept of validity, the definition of measurement in psychology, and the relation between different test theoretic models. He has published on these topics in journals such as *Psychological Review*, *Psychological Methods*, and *Psychometrika*. Current research focuses on the relation between complex systems theory and psychometric models, the application of graphical modeling to psychopathology, and the analysis of test validity.